

Wind Power System Risk Assessment Based on Fuzzy Clustering and Copula Function Modeling

Mingshun Liu¹, Lijin Zhao¹, Liang Huang¹, Wenhao Han², Changhong Deng², Zhijun Long²

¹Guizhou Power Grid Co., Ltd., Guiyang, China

²School of Electrical Engineering, Wuhan University, Wuhan, China

Email: hanwenhao3258@163.com

How to cite this paper: Liu, M.S., Zhao, L.J., Huang, L., Han, W.H., Deng, C.H. and Long, Z.J. (2017) Wind Power System Risk Assessment Based on Fuzzy Clustering and Copula Function Modeling. *Energy and Power Engineering*, 9, 352-364.
<https://doi.org/10.4236/epe.2017.94B041>

Received: February 26, 2017

Accepted: March 30, 2017

Published: April 6, 2017

Abstract

According to the characteristics of the correlation of multiple wind farm output, this paper put forwards a modeling method based on fuzzy c-means clustering and the copula function, and correlation wind farms are inserted into IEEE-RTS79 reliability system for risk assessment. By the probabilistic load flow calculated by Monte Carlo simulation method, the probability of the accident is derived, and bus voltage and branch power flow overload risk index are defined in this paper. The results show that this method can realize the modeling of the correlation of wind power output, and the risk index can identify the weakness of the system, which can provide reference for the operation and maintenance personnel.

Keywords

Correlation, Fuzzy Clustering, Copula Function, Risk Assessment

1. Introduction

Safety is the key of the power system. With the development of wind power technology and large-scale wind power integration, the strong stochastic volatility is bound to bring more serious challenges to stable operation of the system [1]. Besides considering that the same area may have multiple wind farms, due to the similarity of factors such as geographical environment, its output will show some kind of relationship [2]. So it's necessary to conduct the risk assessment of electric power system considering wind power correlation to identify the system weak link, and then take the corresponding effective measures to ensure safety and steady operation.

To consider output correlation of wind power and then conduct risk assessment, modeling the correlation problem is the beginning. Copula function [3] is effective in correlation problem. [4] connects copula theory with Monte Carlo

simulation method for probabilistic load flow calculation. [5] [6] use a hybrid copula function for the modeling of input variables of correlation, and determined the weighting coefficient of each copula function through expectation maximization method and least square method, overcoming the deficiency of using only one copula function. In [8], the author established wind farm reliability model considering the influence of uncertain factors, then proposed a risk assessment method for the composite generation and transmission systems including wind farms based on dispersed sampling Monte Carlo algorithm. [9] establish adequacy evaluation model for composite generation and transmission systems which contain wind farms based on sequential Monte Carlo simulation method. Based on vulnerability of the risk theory evaluation system, the consequences severity with linear function was quantified in [10], but shelter phenomenon exists. In [11], utility function was introduced to measure severity of consequences caused by element fault for the failure probability model of the overhead line.

In this paper, the fuzzy C means clustering is applied to wind power output data firstly and copula function is modeled for each class. The probabilistic load flow of wind power is calculated by Monte Carlo simulation, so the probability of the accident was derived. The utility function and the risk theory is combined to quantify the risk indicators. Matlab simulation results show that the method can assess system risk accurately, and identify system weaknesses, which has significance for power system planning operation, differentiation operation and maintenance.

2. Wind Power Output Correlation Modeling

2.1. Copula Function Theory

Copula can joint distribution of multidimensional random variables with one-dimensional marginal. Take binary random variable as an example to introduce copula function.

$H(x, y)$ is a two joint distribution function with the edge distribution $F(x)$ and $G(y)$, Sklar Theorem points out that there exists the unique Copula function $C(U, V)$ which meets:

$$H(x, y) = C[F(x), G(y)] \quad (1)$$

Copula function mainly include normal copula function and t-Copula function which belong to ellipsoidal copula function, and the Clayton Copula function, Gumbel Copula function, which are the memberships of Archimedes Copula functions. There are differences among different copula functions when they describe the correlation between random variables. Normal copula function, t-Copula function and Frank Copula function are effective in describing the dependence structure of symmetry. While the Clayton Copula function and Gumbel Copula function are used to describe dependence structure of asymmetric, one describes the strong upper tails correlation of the random variables and the other describes the lower tails. In order to describe the correlation between random variables quantitatively and accurately, the results are usually compared

with empirical Copula distribution functions so as to select the optimal Copula. The empirical Copula function is defined as follows:

(X_p, Y_j) ($i = 1, 2, \dots, n$) is samples form bivariate population (X, Y) . The empirical distribution functions of X and Y are $F_n(x)$ and $G_n(y)$ respectively, the sample empirical Copula distribution function was:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n I_{[F_n(x_i) \leq u]} I_{[G_n(y_i) \leq v]} u, v \in [0, 1] \tag{2}$$

where $I_{[\cdot]}$ is an indicative function. It means when $F_n(x_i) \leq u$, $I_{[F_n(x_i) \leq u]} = 1$; otherwise, $I_{[F_n(x_i) \leq u]} = 0$.

Through calculating and comparing square Euclidean distance of each Copula function and empirical Copula distribution function, optimal function can be obtained.

$$d_m^2 = \sum_{i=1}^n |C_n(u, v) - C_m(u, v)|^2 \tag{3}$$

where m is the chosen Copula function type, $C_n(u, v)$ is empirical Copula distribution function, $C_m(u, v)$ is the selected Copula distribution function, d_m^2 is the square Euclidean distance. The smaller value shows that the selected Copula function is more effective in depicting correlation. In this paper, squared Euclidean distance is used to quantified the correlation degree of Copula function.

2.2. Fuzzy Clustering

The final clustering results of traditional clustering algorithms such as K-means depends on the choice of initial aggregation point or the number of strict classification in some degree. While fuzzy clustering aims at the optimization of the objective function, dynamically adjusts the clustering center and the membership degree, and then determines the class of the sample points by iterative convergence so as to automatically classify the sample data. In this paper, the fuzzy C means clustering is used for the wind farm output classification.

$$X = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \tag{4}$$

X is a given sample matrix, p is the number of random variables, n is the number of random variables. Fuzzy clustering is to divide the n observations into c class, the clustering center is $V = \{v_1, v_2, \dots, v_c\}$, of which $v_i = (v_{i1}, v_{i2}, \dots, v_{ip})$ ($i = 1, 2, \dots, c$).

u_{ik} is the membership grade of class i membership, and $\sum_{i=1}^c u_{ik} = 1$, the objective function is defined as:

$$J(U, V) = \min \left\{ \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \| \mathbf{x}_k - \mathbf{v}_i \|^2 \right\} \tag{5}$$

$\mathbf{U} = (u_{ik})_{c \times n}$ shows membership matrix $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|$. The objective function value $J(\mathbf{U}, \mathbf{V})$ is expressed as the weighted square distance and the weighted square distance between the sample and the cluster center.

The specific steps are:

1) Determine the c number of classes, power exponent m and the initial membership matrix $\mathbf{U}^{(0)} = (u_{ik}^{(0)})$, determine the initial membership matrix $\mathbf{U}^{(0)}$ through a series of random numbers produced by a uniform distribution in $[0,1]$.

2) l is iteration step number. The cluster center at step l is:

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^n (u_{ik}^{(l-1)})^m \mathbf{x}_k}{\sum_{k=1}^n (u_{ik}^{(l-1)})^m} \quad (i = 1, 2, \dots, c) \quad (6)$$

3) Modify membership matrix $\mathbf{U}^{(l)}$, then calculate the value of the objective function $J^{(l)}$.

$$u_{ik}^{(l)} = 1 / \sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|^{(l)}}{\|\mathbf{x}_k - \mathbf{v}_j\|^{(l)}} \right)^{\frac{2}{m-1}} \quad (i = 1, 2, \dots, c; k = 1, 2, \dots, n) \quad (7)$$

$$J^{(l)}(\mathbf{U}^{(l)}, \mathbf{V}^{(l)}) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik}^{(l)})^m (\|\mathbf{x}_k - \mathbf{v}_i^{(l)}\|)^2 \quad (8)$$

4) Determining membership tolerance of terminating iteration $\varepsilon_u > 0$. When $\max\{|u_{ik}^{(l)} - u_{ik}^{(l-1)}|\} < \varepsilon_u$, the iteration terminates, otherwise $l = l + 1$, then turn to step (2).

Through the above steps, the final cluster center \mathbf{V} and the membership \mathbf{U} can be obtained, and sample class can be determined according to the element value of \mathbf{U} .

3. Power System Risk Index

Power system risk is a comprehensive measurement system of probability and the seriousness of the consequences of the accident [12], which can reflect the effects of the accident on the operation, according to the theory of risk, the risk can be expressed as a product of the accident probability and severity, expression is as follows [13],

$$Risk = Pro \cdot Sev \quad (9)$$

where $Risk$ is risk value, Pro is the probability of accident, Sev is the severity of accident consequence.

The severity of accident consequence is described by degree of deviation between actual value and rated value. This paper uses risk utility function to describe severity, w is risk index, S is utility function value, $S'(w) > 0$, $S''(w) > 0$. These means with the increase of deviation degree, the speed of the serious increase also accelerated, which is close to the actual operation of the power system. With the tendency of wind power and other new energy sources are integrated into grid, the maintenance of voltage level and the ability to withstand high power are of great significance to the stable operation of the system. In or-

der to master the security of power system, this paper defines voltage over limit risk and branch flow overload risk index.

3.1. Voltage over Limit Risk

The voltage over limit risk describes the possibility and harm degree of the node voltage limit in the system, which reflects the risk of voltage collapse when the voltage value deviates from the normal operating level. The magnitude of the voltage determines the severity of the voltage over limit, and the severity is quantified by the deviation between the actual value and the rated value. The node voltage 1.0 pu means the severity function value is 0; with the voltage value deviates from the rated value, the severity increases. The node voltage over limit severity function is expressed as:

$$S_{V_i} = e^{L_{LV_i}} - 1, L_{LV_i} = |1 - V_i| \tag{10}$$

$$R_V = \sum_{i=1}^{N_V} P_{V_i} \cdot \alpha_i S_{V_i} \tag{11}$$

where S_{V_i} is voltage node i over limit severity, V_i is the voltage, L_{LV_i} is the voltage fluctuation deviation; R_V is the system voltage over limit the total risk, P_{V_i} is the probability of node i voltage over limit, α_i is the weight factor, N_V is node number.

3.2. Branch Power Flow Overload Risk

Transmission line has transmission power limit, branch flow overload risk reflects the line withstand certain transmission power possibility and harm degree. In order to avoid the occurrence of masking phenomenon, but not ignore the potential risks which line is close to limit completely, risk appears when the line load rate reach 90%. The branch power flow overload severity function is defined as:

$$S_{L_i} = e^{L_{oi}} - 1, L_{oi} = I_i - 0.9L_i \tag{13}$$

$$R_L = \sum_{i=1}^{N_L} P_{L_i} \cdot \beta_i S_{L_i} \tag{14}$$

where S_{L_i} is a branch of I power flow overload severity, I_i is the current trend of I value, L_i is power transmission limit of I branch, L_o is power flow deviation; R_L is the total risk system of branch power flow overload, P_{L_i} is the probability of branch i overload, β_i is an important weight factor, N_L is the total branch number.

4. Risk Assessment of Wind Power Access to Power System

Power system risk value can be obtained from probability value and consequence severity. The utility function above can be used to quantify severity. Because of the stochastic fluctuation of wind power, the probability value is obtained by probabilistic power flow calculation [14]. Monte Carlo simulation method [15], widely used in power system, is accurate in calculating probabilistic power flow. In this paper, probabilistic power flow is calculated by Monte

Carlo simulation, then the probability of the bus voltage limit and the power flow overload can be obtained. Concrete steps are:

- 1) Pretreat wind farm raw data and perform fuzzy clustering.
- 2) The edge distribution function is obtained by the kernel density estimation based on nonparametric estimation. Draw edge distribution histogram to observe the input variable dependent structure.
- 3) Calculate the square Euclidean distance for each kind of data to select optimal Copula function to produce the correlated output samples.
- 4) Model power system with wind farm integration; Calculate probabilistic power flow to obtain probability of bus voltage limit and branch flow overload.
- 5) Calculate the node voltage over limit and branch power flow overload severity degree of the system, and define the comprehensive severity as the arithmetic mean of the total severity of N times power flow calculation.
- 6) Multiply the probability with the consequence severity to obtain the risk value.

Figure 1 is the flow chart.

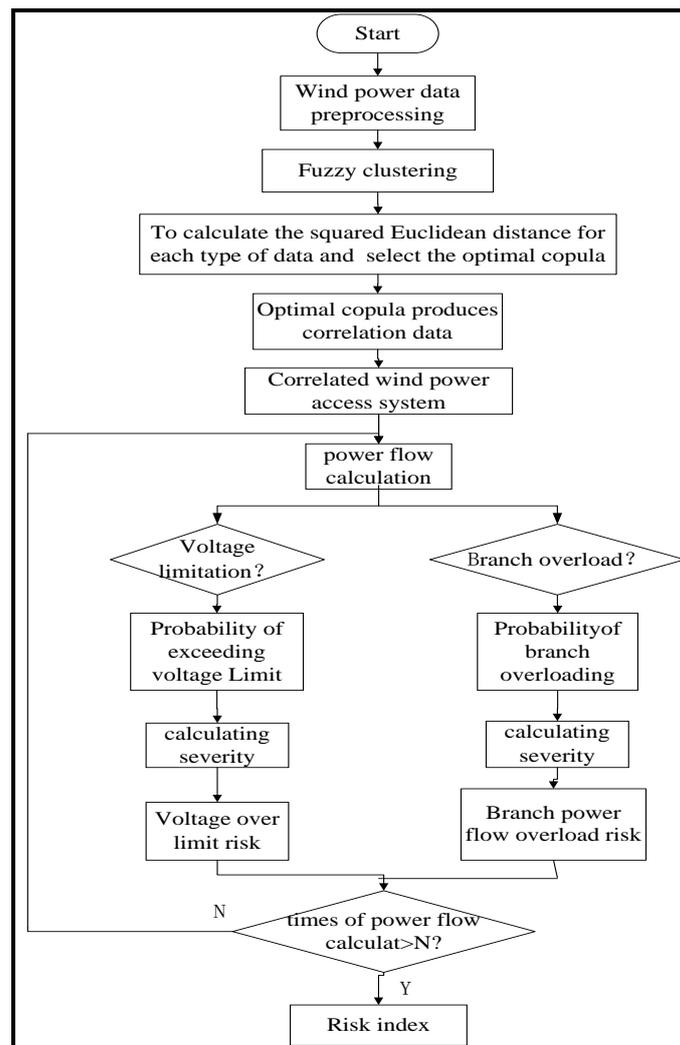


Figure 1. Flow chart of risk assessment.

5. Simulation Results and Analysis

5.1. Wind Power Output Correlation Modeling and Evaluation

Based on the 50,000 sets of measured output data of Australian wind farms in spring, this paper uses fuzzy clustering method combined with Copula function for correlation modeling. The validity of the method is verified by comparing with the measured data.

Fuzzy clustering of the sample matrix is divided into 6 classes; **Table 1** shows the cluster analysis results and the selected optimal Copula function.

To eliminate the influence of sample size difference on correlation coefficient, the total size of the generated data and the measured data should be the same, and produce the output sample of corresponding proportion. **Figure 2** is comparison of the frequency histogram of the measured output and the simulated output.

In **Figure 2**, the left one is frequency histogram of measured output data, the right one is frequency histogram of output by clustering and Copula function. As can be seen from **Figure 2**, there exists correlation between the output of the two wind farms, the correlation is different in different locations, the specific performance of the lower tail has strong correlation, the upper tail is relatively weak, the weakest correlation is in the central. After fuzzy clustering analysis, the lower tail correlation of class 1 and class 4 is depicted by Clayton-Copula function, The Frank-Copula function depicts the symmetry of the other classes. Clustering refines the modeling process and has better fitting effect. The Pearson linear correlation coefficient, Kendall rank correlation coefficient, Spearman rank correlation coefficient and relative error of quantitative σ_i are calculated to analyses excellence of modeling.

$$\sigma_i = \left| \frac{P_{simu} - P_{real}}{P_{real}} \right| \quad (14)$$

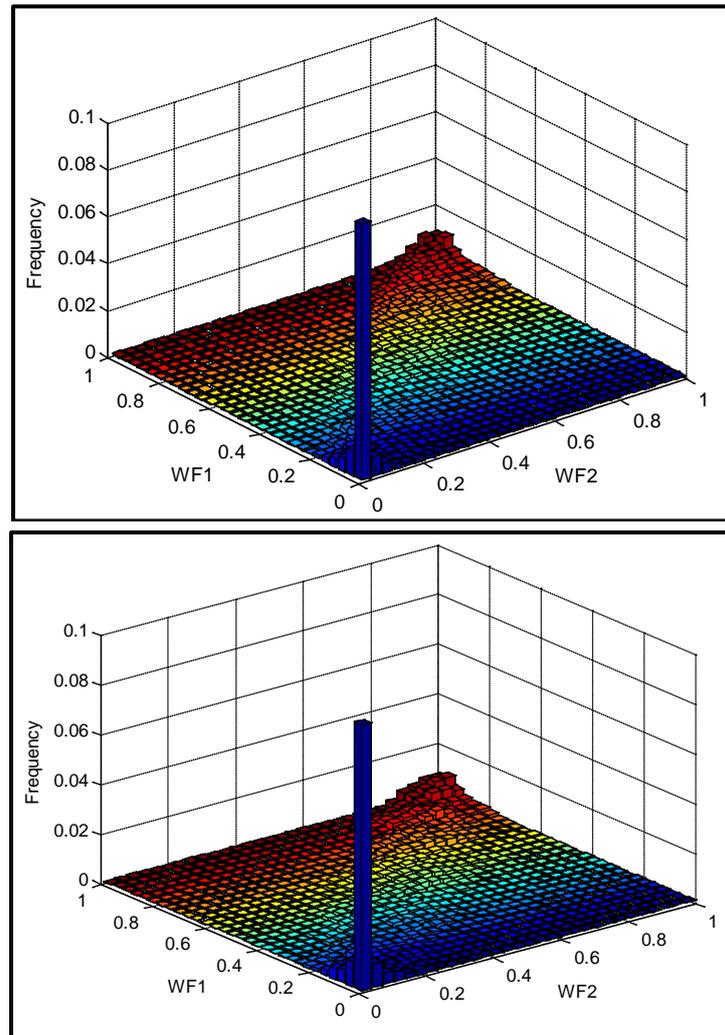
where P_{real} is the wind farm i measured output. P_{simu} is wind farm i simulation output. N represents the total number of samples. For each clustering scheme, the simulated 20 times average is used to reduce the randomness error. **Table 2** is the comparison result of correlation coefficient and relative error between the method and the measured data.

Table 1. Fuzzy clustering results.

Class number	Cluster center	Proportion	Optimal Copula	Squared Euclidean distance
1	(2.2404, 3.5798)	30.81%	Clayton	0.8137
2	(10.4320, 21.5651)	21.71%	Frank	0.6860
3	(20.5277, 43.0317)	16.71%	Frank	0.7204
4	(33.6675, 69.1418)	11.14%	Clayton	0.8094
5	(48.8864, 97.3516)	10.07%	Frank	1.5543
6	(67.9871, 119.5788)	9.56%	Frank	0.4795

Table 2. Correlation and error comparison.

	Pearson	Kendall	Spearman	$\sum_i^N \sigma_i$
Actual Data	0.9140	0.7915	0.9348	-
Copula	0.9221	0.7957	0.9441	0.1976

**Figure 2.** Comparison of measured and simulated output data.

From the above table, we can see that all kinds of optimal Copula functions generated by fuzzy clustering are clustered around the center of clustering, the concentration is strong and the relative error is smaller. This method can accurately describe the correlation of wind power output.

5.2. Risk Assessment of Wind Power Access to Power System

The above two wind farms are respectively integrated into IEEE-RTS79 reliability test system node 17 and 24, wind turbine takes constant power factor control method, and its power factor is $\cos\varphi = -0.95$, integration node is taken as PQ node with negative power and simulation scale $N = 50,000$. The load fluctuation

is random variable which obeys normal distribution, the expectation is the given value of the standard system, and variance is 5% of the expectation, **Figure 3** is the example.

When the two correlation wind farms are integrated into node 17 and 24, the voltage fluctuation increases at the access point, the voltage shows a downward trend, and the low voltage over limit may appear. **Table 3** is partial node voltage information. The voltage fluctuations before and after wind power integration are shown in **Figure 4** and **Figure 5**.

Compared with the no integration only considering the random fluctuation of load, the volatility of node far away from the integration node (such as node 6) is slightly enhanced, but still fluctuates in the safe range; Node 17 and 24 voltage and access node nearby (such as node 3) showed larger variance, the voltage fluctuation is greatly increased, the minimum voltage is lower than the lower

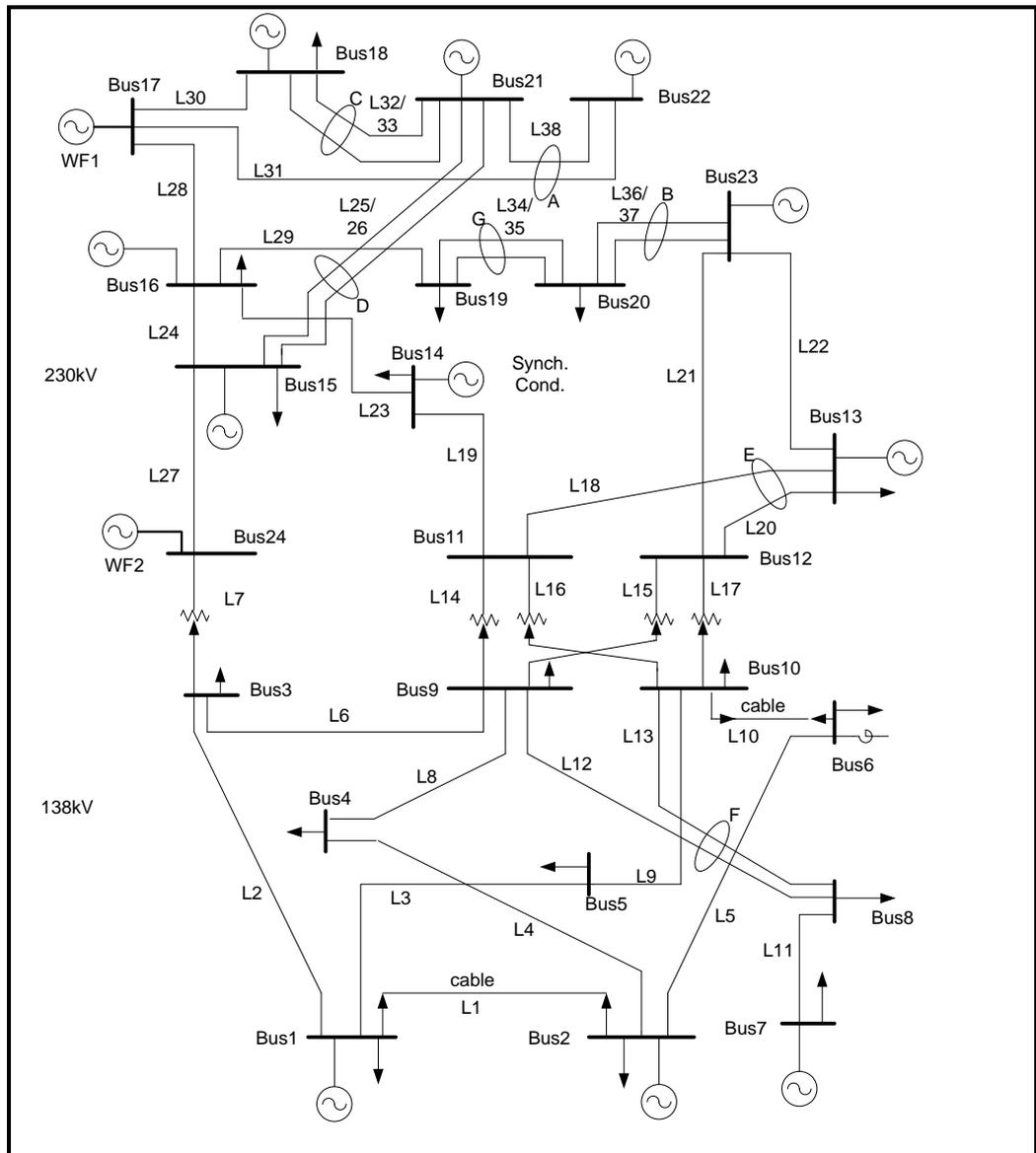


Figure 3. IEEE-RTS79 system with wind power integration.

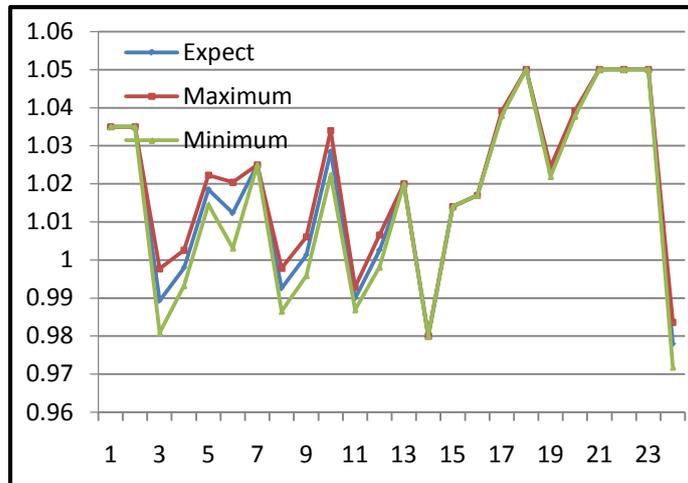


Figure 4. Voltage without wind integration.

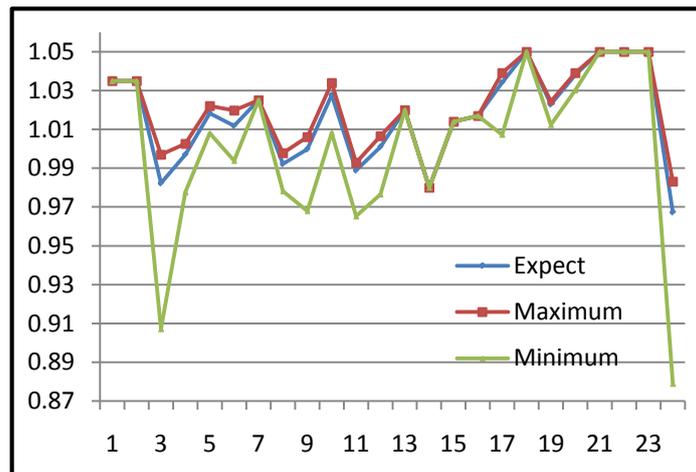


Figure 5. Voltage with wind integration.

Table 3. Node voltage information.

Node	Expect	Maximum	Minimum	Variance
3	0.983616/	0.997678/	0.980912/	3.9188e-06/
	0.982288	0.997068	0.907201	3.7163e-04
6	1.012343/	1.020438/	1.003119/	3.9139e-06/
	1.011934	1.019752	0.993875	1.12633e-05
17	1.038548/	1.039121/	1.037869/	2.2149e-08/
	1.034183	1.039034	1.007277	5.4045e-05
24	0.977827/	0.983634/	0.971891/	1.9595e-06
	0.967348	0.983138	0.878941	5.5165e-04

Notes: “*/” in the table means “no wind integration/wind integration” data.

limit, there is a low voltage situation and a voltage over limit risk. Wind power access changes the original power flow distribution, so the branch power flow may reach the transmission limit of line, destroy the thermal stability of the line and cause overload phenomenon, which leads to the fault of relay protection

operation and may cause cascading failures also if serious. According to risk theory formula, the importance factor is taken as 1, the system node voltage over limit and branch flow overload risk are calculated as shown in **Table 4**.

In order to quantitatively characterize the line carrying capacity when the maximum power flow occurs, the maximum load rate is defined:

$$\lambda_m = \frac{S_{max}}{S_{lim}} \tag{13}$$

where s_{max} is the branch maximum power flow, S_{lim} is the maximum transmission limit.

Figure 6 is the maximum load overload branch flow rate, the penetration of wind power for η , pictures from left to right respectively shows the maximum load ratio of without wind, permeability of 0.4η and η .

Branch 10 has been close to full load while it is not connected to the maximum load rate of wind power, therefore, the line transmission capacity should be increase to reduce risk. The power flow of branch 18 and branch 27 has changed greatly after the access of wind power. With the increase of permeability, the maximum line load rate increases gradually, and the grid risk exists. Branch 18 is the transformer branch, branch 24 is a high voltage class 230 kV, and is an important channel for the transmission of electricity to the 230 kV

Table 4. Risk value.

Node/Branch	Probability	Severity	Risk
Node 3	13.528%	0.0261	0.003531
Node 24	25.946%	0.0431	0.011183
Branch 10	69.596%	0.2125	0.147891
Branch 18	5.592%	0.0906	0.005066
Branch 27	1.5%	0.0318	0.000470

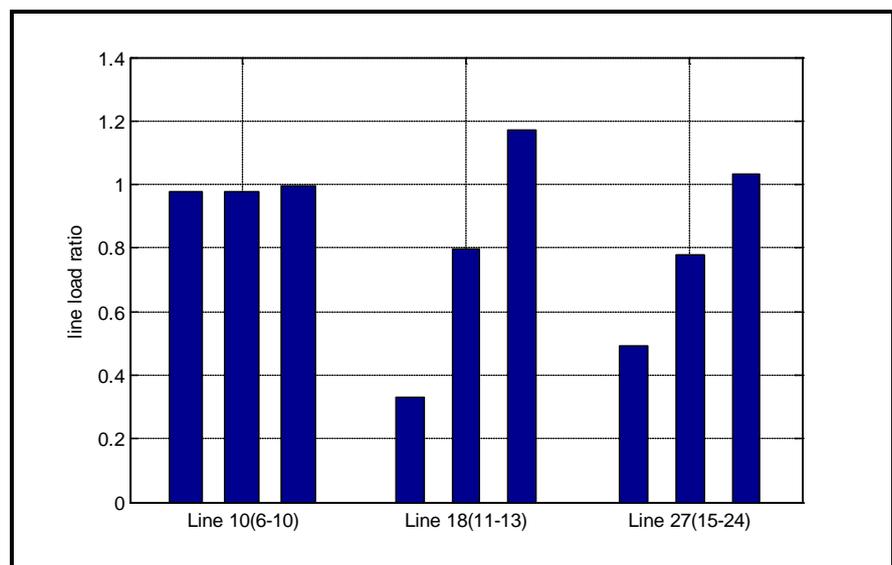


Figure 6. Overload current maximum load rate.

area, there should be attention to the risk.

As can be seen, in the current wind power access mode, node 3 and 24 have the risk of low voltage, 10, 18 and 27 have the branch power flow overload risk, which can be regarded as the key nodes and lines to be paid attention to. The system personnel should carry on the pertinence analysis, reasonably plan the wind power access point and the access capacity, take the corresponding measure to reduce the electric network risk, and provides the safeguard for the power system safe reliable operation.

6. Conclusions

Based on the measured output data of wind farm, the fuzzy clustering and Copula function theory are combined to realize the correlation modeling of wind power output. The wind power probabilistic flow is calculated by Monte Carlo simulation method to obtain the probability of the node voltage over limit and branch flow overload. The severity is measured by utility function, and the risk value is calculated according to the risk theory. Results show that:

- 1) After the fuzzy clustering processing for the total sample, the optimal Copula function is modeled for each kind of data, which can accurately describe the correlation of wind power output.
- 2) In and near the wind power access position, the voltage fluctuation is strong and prone to have voltage over limit risk.
- 3) This method can evaluate the risk of branch flow overload, identify the system critical lines and provide theoretical support for differential operation and maintenance.

This method in the paper can realize the modeling of wind powers with correlation, quantitative risk index of limit over voltage and branch flow overload and realize the power grid risk assessment of the operation condition. It can identify the weak links and key lines when the wind power access to the power system which can provide the basis for the realization of the difference operation.

References

- [1] Bart, C.U. and Madeleine G, Wil L K. (2007) Impacts of windpower on thermal generation unit commitment and dispatch. *IEEE Transactions on Energy Conversion*, **22**, 44-51. <https://doi.org/10.1109/TEC.2006.889616>
- [2] Rong X.X., IE Z.H., Shi, W.H., et al. (2014) Analysis on Probabilistic Load Flow in Power Grid Integrated With Wind Farms Considering Correlativity Among Different Wind Farms. *Power System Technology*, **38**, 2161-2167
- [3] Pagaefthymiou G. (2009) Using Copulas for modeling stochastic dependence in power system uncertainty analysis. *IEEE Trans on Power System*, **4**, 40-49. <https://doi.org/10.1109/TPWRS.2008.2004728>
- [4] Cai, D.F., Shi, D.Y. and Chen, J.F. (2013) Probabilistic load flow considering correlation between input random variables based on Copula theory. *Power System Protection and Control*, **41**, 13-19.
- [5] Pan, X., Wang, L.L., Xu, Y.Q., et al. (2014) Wind Power Correlation Analysis Based

- on Hybrid Copula. *Automation of Electric Power Systems*, 2014, **38**, 17-22.
- [6] Ji, F., Cai, X.G. and Wang, J. (2014) Wind Power Correlation Analysis Based on Hybrid Copula. *Automation of Electric Power Systems*, **38**, 1-5+32.
- [7] Wang, J., Cai, X.G., and Ji, F.(2013) A Simulation Method of Correlated Random Variables Based on Copula. *Proceedings of the CSEE*, **33**, 75-82+13.
- [8] Jiang, C., Liu, W.X., Zhang, J.H., et al. (2014) Risk Assessment of Generation and Transmission Systems Considering Wind Power Penetration. *Transactions of China Electrotechnical Society*, **29**, 260-270.
- [9] Zhang S., Li, G.Y. and Zhou, M. (2010) Reliability Assessment of Generation and Transmission Systems Integrated With Wind Farms. *Proceedings of the CSEE*, **30**,8-14.
- [10] Chen, W.H., Jiang, Q.Y., Cao, Y.J., et al. (2005) Risk-Based Vulnerability Assessment in Complex Power Systems. *Power System Technology*, **29**, 12-17.
- [11] Zhang, Y.M., Zhang, Z.H., Yao, F., et al. (2013) Risk assessment of power system components based on the risk theory. *Power System Protection and Control*, **41**, 73-78.
- [12] (1997) GIGRE Task Force 38.03.12. Power System Security Assessment, a Position Paper, Elctra, 1997, 175: 49-77
- [13] Fu, W.H. andMcCalley, J.D. (2001) Risk Based Optimal Power Flow.*IEEE Power Tech Proceedings*, Porto, Portugal, 2001.
- [14] Dong, L., Cheng, W.D. and Yang, Y.H. (2009) Probabilistic Load Flow Calculation for Power Grid Containing Wind Farms. *Power System Technology*, **33**, 87-91.
- [15] Billintonr, L.Y. (1994) Reliability Evaluation of Electric Power Systems Using Monte Carlomethods.Plenu Press,New York.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact epe@scirp.org